

REPORT DOCUMENTATION PAGE

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing existing data, gathering the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE (DD-MM-YYYY) 11/30/09		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 07/15/09-11/30/09	
4. TITLE AND SUBTITLE Paths Ahead in the Science of Information and Decision Systems - Workshop				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER FA9550-09-1-0498	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Prof. Munther Dahleh				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Massachusetts Institute of Technology 77 Massachusetts Ave Cambridge, MA 02139				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR 875 N Randolph St., Ste 325 Arlington, VA 22203 RSL				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution A: Approved for Public Release					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT See report attached.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)

20100622250

**Laboratory for Information and Decision Systems (LIDS)
Massachusetts Institute of Technology
Cambridge, MA 02139**

**Final Technical Report for
Grant FA9550-09-1-0498**

Paths Ahead in the Science of Information and Decision Systems – Workshop

Principal Investigator: Prof. Munther Dahlah

Submitted to: Dr. Robert Bonneau
AFOSR-OSR
875 N. Randolph Street, Room 3112
Arlington, VA 22203

I. Abstract

A significant meeting, the symposium on “Paths Ahead in the Science of Information and Decision Systems” was held November 12-14, 2009 at MIT. This meeting was organized and run by MIT’s Laboratory for Information and Decision Systems (LIDS), the oldest continuing laboratory at MIT. LIDS has played and continues to play a major role in the development of our field, responding to critical national and societal needs; developing fundamental and path-breaking advances in theory, methodology, and practice; and serving as a focal point for activities involving the best across MIT, the nation, and the world.

The science of information and decision systems encompasses a substantial and exceptionally pervasive set of interrelated disciplines, ranging from signal and image processing; to embedded control systems; to the analysis, design, and optimization of complex distributed systems and networks. Thanks both to the richness of the challenges throughout engineering and the physical, biological and social sciences, and the continuing developments of the foundations of our disciplines, the information and decision sciences stand today as an exciting, continually evolving, and critical domain of intellectual inquiry.

Consistent with that history and mission, LIDS organized the Paths Ahead Symposium, bringing together leading researchers from all around the world who have been influential in shaping the vision of and leading this broad field. The meeting, which was sponsored by MIT’s School of Engineering, by a number of private companies and laboratories, and by NSF, AFOSR, and ARO, consisted of several panel-oriented sessions, providing both context and history as well views toward the challenges of the future. While each of these sessions had a specific theme, an overall objective of each session was to look across disciplines for challenges and opportunities across disciplines.

II. Panels Detailed Description

The meeting, which attracted 340 registrants, began with a reception on Nov. 12 at the MIT Museum, and the technical sessions began on Friday, Nov. 13 with welcoming remarks by Symposium General Chair and LIDS Director, **Alan Willsky**.

The morning session on Friday was organized in honor of **Sanjoy Mitter**, a major leader in the field and former Director of LIDS, who recently retired from MIT (although one would not know that from his continuing presence). This session was chaired by **Thomas Magnanti** (MIT), former Dean of Engineering at MIT and long-time collaborator with Sanjoy, and the panelists were **Karl Johan Åström** (Lund Inst.), **Dimitri Bertsekas** (MIT), **Roger Brockett** (Harvard), **Y.C. (Larry) Ho** (Harvard), **Thomas Kailath** (Stanford), **Petar Kokotovic** (UC Santa Barbara), **Pravin Varaiya** (UC Berkeley), and **Jan Willems** (Cath. U. of Leuven). The presentations and discussion in this section ranged from personal perspectives on the past history of research in this broad field, on some of the challenges and exciting opportunities that are before us, and on the challenge of educating our students in a field with the breadth that the information and decision sciences possess. Among the challenges for the future brought up by the panel were those central to the social agenda and the exciting areas of research of today, including energy, transportation, biology, and health care. Intellectual challenges including

developing new methods to deal with compositional descriptions of complex systems and the broad area of networks, information, and control were also discussed, as were continuing areas such as robotics, embedded systems, and autonomy. The intellectual vibrancy of our field and its ability to move with agility into new domains, were evident throughout.

The Friday afternoon session had as its “center of gravity” systems, control, and optimization. The session was chaired by LIDS Co-Associate Director, **Munther Dahleh** (MIT), with a panel consisting of principal speaker **Keith Glover** (Cambridge) and panelists **Albert Benveniste** (INRIA), **Vincent Blondel** (Cath. U. of Louvain), **Stephen Boyd** (Stanford), **Jonathan How** (MIT), **Richard Murray** (Cal. Tech.), and **Pablo Parrilo** (MIT). Keith Glover gave an overview of the field and its central elements, including feedback, dealing with uncertainty, approximation, and verification and certification of performance, as well as a discussion of areas appropriate for academic research, ranging from design methodologies for particular applications to development of verification tools. The other presentations in the session spanned topics including a discussion of the central role that computational methods play in our field (and in particular in redefining what we mean by a “solution”); a presentation of grand challenges (including robust and certifiably correct control of networked systems such as smart grids, the need for learning algorithms that lead to safe performance in a nonstationary world, and NASA’s Green Flight Challenge; a discussion of the challenges in controlling complex systems, with examples including the DARPA Urban Challenge and the stunningly robust, computationally limited, and slow control system that allows a fly to maintain stable flight in the presence of sudden changes such as wind gusts; and a presentation on the challenges of “componentizing” control systems as is generally specified in the system engineering of complex and often safety-critical man-made systems and the clear need for those in systems and control to contribute to the overall system-wide issues as well as to the components.

The Symposium banquet was held Friday evening and included a banquet talk by **Alan Willsky** on the long and celebrated history of LIDS, beginning with its days as the Servomechanism Laboratory extending back to the period prior to the Second World War up through the present. The talk presented a picture of the major figures whose contributions fueled the Laboratory’s major role in academia and society, as well as highlighting many of the contributions over the years including the development of high-performance fire control systems, the invention of magnetic core memory, major advances in numerically controlled machines, some of the earliest efforts in CAD and database systems, a leadership role in the development of modern control and the development of robust control methods, the broadening of its agenda to include networked systems ranging from communication and transportation networks to power grids as well as the expansion of efforts in statistical signal processing and learning, continuing advances in optimization algorithms, and through this entire history, a strong record of theoretical advances, influential texts, and an impressive array of former students, colleagues, and visitors.

Saturday morning, Nov. 14, began with a plenary talk by **Sanjoy Mitter**, entitled “System Theory: A Retrospective and Prospective Look.” This far-reaching lecture provided both concrete and philosophical remarks about revolutionary science and argued that such a revolution took place in system theory in the 1960’s, with its key elements being the emerging central role of computation, a new language leading to state space models, and the exploitation

of this language to gain a far deeper understanding of systems as well as powerful new methods. Mitter argued that the challenges of today, in particular networked systems, might require some new elements and lead to new structural insights and methods. The talk also touched on pattern recognition and artificial intelligence and their close intellectual ties to information and decision systems as well as ties of Bayesian inference to statistical mechanics, a topic that resonates with the role of physics in understanding some of the core models and methods in machine learning. Mitter also discussed the challenges and opportunities that arise when one brings the constraints of communication systems into the design of control systems and closed with a list of challenges that could, by itself, fuel the field for a very long time.

The Saturday morning panel discussion had networks and information, broadly defined, as its center of gravity. This session was chaired by LIDS Co-Associate Director, **John Tsitsiklis** (MIT). **John Doyle** (Cal. Tech.) was the lead speaker in the session, together with a panel consisting of **P.R. Kumar** (Illinois), **Asuman Ozdaglar** (MIT), **H. Vincent Poor** (Princeton), **Balaji Prabhakar** (Stanford), **Jeff Shamma** (Georgia Tech.), and **David Tse** (UC Berkeley). In his presentation John Doyle gave a far-reaching discourse on networks, layered systems, their fragility and challenges in their design, as well as a contrasting view of some man-made networks (the Internet, power grids, etc.) and biological systems (e.g., bacteria), pointing out similarities, differences, and challenges for those of us in the information and decision sciences. The presentations of other panelists included: an examination of application and domain challenges (including wireless security and multimedia communications) to “pull” the development of methodology and the “push” of specific technical challenges (e.g., in information theory and finite-blocklength capacity); an examination, through example, of why it is worthwhile to continue examining very hard problems and looking for ways in which to reformulate them creatively in ways that overcome technical difficulties and lead to new results and insights; an examination of the serious challenges in interplay of networks and information (including control of distributed systems over unreliable networks, methods for verifying performance, and distributing information processing as a problem blending computation, communication, and inference); the design of incentive systems for complex transportation networks in order to influence behavior and reduce congestion; challenges and opportunities in network games and in understanding dynamics, learning, and decision-making in social and economic networks.

The Saturday afternoon panel discussion, focusing on signal processing, inference, and learning, was chaired by **Alan Willsky**. The lead speaker in this session was **Michael Jordan** (UC Berkeley), who was joined on the panel by **Alfred Hero** (Michigan), **Sanjeev Kulkarni** (Princeton), **Robert Nowak** (Wisconsin), **Pietro Perona** (Cal. Tech.), **Devavrat Shah** (MIT), and **Martin Wainwright** (UC Berkeley). Mike Jordan’s presentation provided an overview of the broad area of machine learning and its ties to problems in a vast array of fields. In this presentation he provided a view of current trends in machine learning including: Nonparametric Bayesian methods (with applications in signal and image processing highlighted), the challenges that the availability of massive data sets presents to those in learning and modeling; the investigation of “Objective Bayes” methods which provide a unifying blend between Bayesian and frequentist views of statistics, with many ties to information theory; the great interest in methods that capture or recover “sparsity” in one form or another; and the challenge of bringing control and statistics together in the same synergistic way as optimization and statistics. Other

presentations provided discussions of machine learning challenges in computer vision (e.g., so that one can search on parts of images or so that we can capture a human's ability to recognize new objects quickly); the challenging dynamic learning problems embedded in the operation of engineered networks (e.g., medium access control) and the role of so-called message passing algorithms; the challenges and opportunities in confronting increasingly high-dimensional data sets (with applications in learning graphical models) and the "blessings" as well as the well-known curses of dimensionality (with applications in sparse reconstruction and the uncovering of scaling laws) as well as the posing of a question seen in other sessions as well, namely the tradeoff between computational effort and performance; the challenges in distributed or networked learning, and the fusion or aggregation of heterogeneous and nontraditional signal and data sources (ranging from sensor outputs to written text to forecasts of multiple agents); integrative modeling, prediction, and uncertainty assessment with predictive health and disease detection as a motivating application and challenge, characterized by heterogeneous data and diverse outputs (ranging from individual predictions to drug effectiveness assessment); and the use of feedback in sensing systems, i.e., the control or selection of measurements to be taken driven by the information state resulting from data already collected.

As indicated, the meeting attracted a substantial number of researchers from around the world, leading to lively discussions that prompted our inviting participants to continue this conversation and to provide short perspective and position papers through the end of 2009. The website for this meeting <http://paths.lids.mit.edu> includes not only a statement of purpose, agenda, list of sponsors, etc., but also a complete collection of files generated by this meeting. This includes (a) video of the entire meeting; (b) all panelist slides; (c) short perspectives and position papers submitted by attendees; and (d) a summary document produced by Munther Dahleh, John Tsitsiklis, and Alan Willsky.

III. Paths Ahead in the Science of Information and Decision Systems

The present section is a summary of some of the key ideas that emerged from the panel discussions and the subsequent exchange of materials. As such, it also serves as a position paper reflecting some of the views of the community on the future directions of the field. A concise description of the topics covered in each presentation and panel discussion can be found in a brief report on this meeting that has been published in several IEEE magazines and newsletters in March and April of 2010. A draft of that article is also available on the Paths Ahead website.

The report is organized along the three broad thematic areas that were covered in the meeting (i) systems, control, and optimization, (ii) networks and information, (iii) signal processing, inference, and learning. On the other hand, a central theme, often reiterated during the meeting, was the close coupling between these different areas. Many of the most interesting current and future research activities are situated right at the interface between different areas.

1. Systems, Control, and Optimization

Historically, the original focus of this field was on the modeling, analysis, and feedback controller design for systems described by linear or nonlinear differential or difference equations, with special emphasis on robustness in the presence of input and model uncertainty. More

recently, numerous challenges have emerged, with the focus shifting towards complex, often distributed, systems. Typical concerns that are driving the field stem from the high-dimensionality of such systems, the simultaneous presence of discrete and continuous dynamics (hybrid systems), and the interaction between physical systems with humans or software.

Some key directions addressed by the panelists

- a) Methodologies for deriving simplified models (**model reduction**) that on the one hand are well-structured and sufficiently low-dimensional to be tractable, while on the other hand, remain faithful to the actual dynamics and capture the essential features of interest. This activity is driven by either an original high-dimensional model, or by raw data (in the latter case, there is a clear synergy with the field of inference). This methodological research proceeds hand in hand with domain-specific work in diverse areas, such as, for example, circuit analysis and design, automotive, space systems, systems biology, animation, etc.
- b) The development of **computational methods** for addressing the analysis and design of feedback control systems has witnessed an enormous growth in the last 10 years. While this remains a vital area of research, the panel emphasized that computational approaches should be geared towards getting more qualitative results and not emphasize the numerical aspect of the solution. Also, the panel addressed interesting future opportunities with respect to real-time, stochastic and robust optimization.
- c) The development of novel, tractable methodologies to deal with nonlinear or **hybrid systems**. Particular emphasis is placed here on algorithmic issues, as well as on **verification methods** that can establish that a given design possesses certain desirable properties. Control theory had traditionally offered rigorous design methods guaranteeing stability and performance of feedback systems. In that context, such methods provided stability and performance certificates that can be used to verify the property of interest. When multiple such systems are connected, providing a performance certificate of the integrated systems remains a challenge and an important area of research.
- d) Methodologies for dealing with **distributed, multi-agent, mobile systems** that achieve desired cooperative behavior through limited information exchange, possibly over an unreliable communication network. Much progress has been made in the area of distributed control over noisy networks; however, the fundamental problems are far from being resolved. Multi-agent mobile systems highlight questions of autonomy as well as the interaction of cyber and physical systems. It is expected that future grand challenges will involve the design of such systems operating in an autonomous or semi-autonomous fashion, incorporating humans in the closed loop. Distributed decision systems are expected to play a critical role in emerging and vital fields such as the management of smart grids with a large number of renewable resources and in the presence of dynamic demands.
- e) Some of the most exciting research in communications and systems and networks cuts across traditional disciplinary boundaries. One such direction involves economic theory

(especially game theory and mechanism design) to (i) study incentive systems that can induce socially desirable behavior on the part of the users, (ii) the effects of different pricing mechanisms, and (iii) the effects of different market structures. Another direction involves the interplay with systems and control theory. On the one hand, control theory provides insights and tools for the **control of networks**; on the other hand, there are several challenges in the field of **control over networks**, whereby feedback controllers operate in the presence of distributed and possibly delayed information that is delivered over a network infrastructure.

- f) While interconnections are fundamental to the design of control systems, much needs to be done to understand **component-based** design. In its simplest form, a theory is needed to understand how specifications on a composite system translate into specifications on components. If such components are subsystems that are built with performance certificates, then an important question is to understand if a specific interconnection can come with performance guarantees which can be established by constructing a performance certificate from those of the subsystems. Such theory of **composition** is also relevant in understanding the interconnections of heterogeneous networks.
- g) **Foundations of Cyber-Physical Systems (CPS)** This area emphasizes the interaction between the actual physical process and the information and decision network layer on top. Modeling the interface of these two layers will play a critical role in many of the applications mentioned before (e.g., smart grid, future transportation). While this interaction can manifest itself in various ways, the following are some critical areas of research.
 - 1. *Verification of hybrid systems.* It is well known that the exact verification problem is computationally intractable, however, there is a need for effective approximation methods that have verifiable properties for certain classes of hybrid systems. This development hinges on the development of tools for analysis and synthesis of classes of hybrid systems along with software verification tools based on discrete verification problems.
 - 2. *Architectures of CPS.* Architecture design of the information layer is central to the design of optimal and robust distributed algorithms that can effectively interact with the physical system. Currently, the theory of interconnections and compositions of subsystems that exhibit a certain performance guarantee is limited to very simple rules that are not general enough to address complex heterogeneous interconnections. Methods for parametrizations of effective architectures and the use of approximation techniques to search among these parametrizations need to be developed to streamline the process of architecture design. In addition, such architectures need to support important attributes such as mobility and reconfigurability.
 - 3. *Cross-Layer algorithms.* In addition to efficient architectures, there is a need to devise algorithms that simultaneously exploit all layers of a system to perform data fusion, communication, computation, and decision making. In a traditional communication

network, this will entail exploiting the interactions between the physical layer, transport layer, and application layer.

2. Networks and Information

The field of information transmission, communications, and networks has a long history and in many respects a mature one. At the same time, related technologies are evolving at a fast pace; many applications appear to be much more complex and multifaceted in comparison with existing analytical models. This apparent complexity motivates a focus on new models and reformulations that will provide new insights and open up new possibilities for progress.

1. Some examples of topics in which new mathematical formulations and methods can have impact are the following
 - a. **Security** is a major concern in modern communication systems. Besides the traditional methods of cryptography (which are based on the presumed intractability of certain computational problems), there is room for secure communication methods that exploit the statistical properties of physical-layer wireless channels.
 - b. While much of information theory has focused on asymptotic formulations, involving large blocklengths, many important applications (multimedia, streaming, etc.) require the use of **small blocklengths**. The problem of calculating approximating capacity and developing good codes subject to short blocklengths and or finite delay constraints becomes an important one.
 - c. **Network information theory** has been slow to develop, despite its great relevance to wireless networks, largely because of the inherent complexity of the underlying problems. One of the lessons of some recent research activities is that it may be fruitful to look for alternative formulations, perhaps involving approximate models, which can deliver qualitative insight, approximate results, and scaling laws, thus bypassing the harder problems of classical network information theory.
 - d. The problem of **medium access control** manifests itself in new ways in the context of dynamically evolving, possibly heterogeneous networks. There is a need for new algorithms that perform well in the presence of uncertainties and time variations (e.g., in the number of competing users), as well as for algorithms that are simple (so that they can be implemented in modern, high-rate networks).
 - e. **Stochastic network theory** remains an important subject, related to optimizing the performance of network schedulers, protocols, and resource allocation mechanisms. It entails the development of novel methods for the analysis of queueing and other types of networks, including model reduction and approximation techniques.
2. There are serious challenges in the **interplay of networks and information** (including control of distributed systems over unreliable networks, methods for verifying performance, and distributing information processing). Such problems typically blend subproblems involving computation, communication, and inference. Issues of synchronization, quality of service guarantees, and distributed real-time scheduling, are

ubiquitous. Furthermore, these problems are tightly linked with the issue of verification and correctness of distributed computing, information processing, inference, or control, operating on top of a networking infrastructure that is subject to noise and failures.

3. In recent years, we have witnessed intense activity related to diverse types of networks (e.g., natural, technological, infrastructure, economic, and social), in multiple disciplines (graph theory, applied probability, statistical physics, economics, game theory, graphical models, sensor networks, biological networks, social networks, multi-agent coordination, etc.). This activity, sometimes loosely referred to as **network science**, is at the gestation stage, with the main intellectual paradigms still under development. Some examples of important problems are as follows.
 - a. Develop an understanding of network performance as a function of high-level qualitative structural properties; for example, the impact of a “small-world” structure on communication network performance.
 - b. Infrastructure networks often involve economic mechanisms, including humans that respond to economic incentives. Thus, an important lever for improving network performance is often the design of proper **incentive schemes**. A prime example is the case of transportation networks, where properly designed monetary incentive mechanisms hold the promise of mitigating the growing problem of resource congestion. In the same spirit, it is important to understand the impact of different market structures, incentive schemes, pricing policies, etc., in the context of communication networks.
 - c. Regarding social networks, there is a need for models that predict the **dynamics of social behavior** on the part of individuals connected through a social network; and, methods for enhancing the robustness of networks involving selfish agents. Going further one step, there are important problems of how to induce desirable behavior on a society of individuals that interact through a network. More generally, many networks are time-varying (in the sense that either the network structure or the agents’ beliefs keep changing), but allow the possibility of affecting the time-evolution through appropriate supervisory action (e.g., to mitigate disturbances or to ensure some desired structural properties). This leads to a new and promising subject of **feedback control of evolving networks**.
4. The problem of **architecting complex networked** systems is a critical one, and arises in the numerous contexts involving infrastructure networks (e.g., communication networks, the power grid, transportation, etc.). The development of new ways of thinking about architectural principles may enable systematic approaches for influencing the development of infrastructure networks (e.g., in architecting the future internet), as opposed to incremental interventions. New approaches may emerge through a thorough understanding of the architectural principles behind biological systems, in order to avoid key drawbacks (such as lack of robustness and fragility) of many man-made complex systems.

3. Signal processing, inference, and learning

This broad and pervasive field encompasses problems ranging from statistical signal processing to the fusion and extraction of information from highly heterogeneous information sources, to learning and fusion in distributed networks. This field makes strong contact with other core disciplines including information theory, optimization, artificial intelligence, statistics, and approximation theory. The following is a brief list of some of the areas of great interest that were discussed during the Symposium.

1. The fields of **machine learning and statistics** are currently very active. There are multiple fronts on which important advances have been made or are being pursued, and which tie to problems in a vast array of fields. Some important active areas include Nonparametric Bayesian methods (with applications in signal and image processing); “Objective Bayes” methods which provide a unifying blend between Bayesian and frequentist views of statistics, with many ties to information theory; methods that capture or recover “sparsity” in one form or another. There is also the challenge of bringing control and statistics together in the same synergistic way as optimization and statistics: one example is the use of feedback in sensing systems, i.e., the control or selection of measurements to be taken driven by the information state resulting from data already collected. Finally, one particular area in which there are major machine learning challenges is computer vision, e.g., for searching on parts of images or in capturing a human’s ability to recognize new objects quickly.

2. An activity of major importance is the development of learning and modeling methods that can deal with **massive and/or high-dimensional data sets**, including methods for learning graphical models and for sparse reconstruction, and an understanding of the tradeoff between computational effort and performance.

3. There are important new problems of inferring the structure and properties of **large networks**. For example, how can one learn global properties of a network, or infer its detailed structure given incomplete information (e.g., traffic traces)? Such problems are compounded in the presence of time-variations of the network structure or utilization. This is a problem involving potentially massive (but still incomplete) data, calling for new methods. Inference in biological networks raises similar issues.

4. Similarly, there are important new problems of inference and learning in networked systems, in a distributed manner. This includes the fusion or aggregation of heterogeneous and nontraditional signal and data sources (ranging from sensor outputs to written text to forecasts of multiple agents); integrative modeling, prediction, and uncertainty assessment with predictive health and disease detection as a motivating application and challenge, characterized by heterogeneous data and diverse outputs (ranging from individual predictions to drug effectiveness assessment).